VIRULENCE GENES AND PROTEINS, AND THEIR USE

Field of the Invention

This invention relates to the identification of virulence genes and proteins, and their use. More particularly, it relates to their use in therapy and in screening for drugs.

Background to the Invention

E. coli is a member of the *Enterobacteriaceae*, or enteric bacteria, which are Gram-negative microorganisms that populate the intestinal tracts of animals. Other members of this bacterial family include *Enterobacter*, *Klebsiella*, *Salmonella*, *Shigella* and *Yersinia*. Although *E. coli* is found normally in the human gastrointestinal tract, it has been implicated in human disease, including septicaemia, meningitis, urinary tract infection, wound infection, abscess formation, peritonitis and cholangitis.

The disease states caused by *E. coli* are dependent upon certain virulence determinants. For example, *E. coli* has been implicated in neonatal meningitis and a major determinant of virulence has been identified as the K1 antigen, which is a homopolymer of sialic acid. The K1 antigen may have a role in avoiding the host's immunological system and preventing phagocytosis.

Summary of the Invention

The present invention is based on the identification of a series of virulence genes in *E. coli* K1, and also related organisms the products of which may be implicated in the pathogenicity of the organism.

According to one aspect of the present invention, a peptide is encoded by an operon including any of the genes identified herein as *mdoG*, *creC*, *recG*, *yggN*, *tatA*, *tatB*, *tatC*, *tatE*, *eck1*, *iroD*, *iroC*, *iroE*, *mtd*2 and *ms*1 to 16, from *E. coli* K1, or a homologue thereof in a Gram-negative bacterium, or a functional fragment thereof. Such a peptide is suitable for therapeutic use, e.g. when isolated.

The term "functional fragments" is used herein to define a part of the gene or peptide which retains similar therapeutic utility as the whole gene or peptide. For example, a functional fragment of the peptide may be used as an antigenic determinant, useful in a vaccine or in the production of antibodies.

A gene fragment may be used to encode the active peptide. Alternatively, the gene fragment may have utility in gene therapy, targetting the wild-type gene *in vivo* to exert a therapeutic effect.

A peptide according to the present invention may comprise any of the amino acid sequences identified herein as SEQ ID NOS. 2, 5, 7, 9, 11, 12, 13, 14, 16, 23, 24, 25, 26, 28, 31, 29, 32 and 35-48.

The identification of these peptides as virulence determinants allows them to be used in a number of ways in the treatment of infection. For example, a host may be transformed to express a peptide according to the invention or modified to disrupt expression of the gene encoding the peptide. A vaccine may also comprise a peptide according to the invention, or the means for its expression, for the treatment of infection. In addition, a vaccine may comprise a microorganism having a virulence gene deletion, wherein the gene encodes a peptide according to the invention.

According to another aspect of the invention, the peptides or genes may be used for screening potential antimicrobial drugs or for the detection of virulence.

A further aspect of this invention is the use of any of the products identified herein, for the treatment or prevention of a condition associated with infection by a Gram-negative bacterium, in particular by *E. coli*.

<u>Description of the Invention</u>

The present invention has made use of signature-tagged mutagenesis (STM) (Hensel *et al*, Science, 1995;269:400-403) to screen *E. coli* K1 strain RS228 (Pluschke *et al*, Infection and Immunity 39:599-608) mini-Tn5 mutant bank for attenuated mutants, to identify virulence genes (and virulence determinants) of *E. coli*.

Although *E. coli* K1 was used as the microorganism to identify the virulence genes, corresponding genes in other enteric bacteria are considered to be within the scope of the present invention. For example, corresponding genes or encoded proteins may be found, based on sequence homology, in *Enterobacter*, *Klebsiella* and other genera implicated in human intestinal disease, including *Salmonella*, *Shigella* and *Yersinia*.

The term "virulence determinant" is used herein to define a product, e.g. a peptide or protein that may have a role in the maintenance of pathogenic bacteria. In particular, a virulence determinant is a bacterial protein or peptide that is implicated in the pathogenicity of the infectious or disease-causing microorganism.

A gene that encodes a virulence determinant may be termed a "virulence gene". Disruption of a virulence gene by way of mutation, deletion or insertion, will result in a reduced level of survival of the bacteria in a host, or a general reduction in the pathogenicity of the microorganism.

Signature-tagged mutagenesis has proved a very useful technique for identifying virulence genes, and their products. The technique relies on the ability of transposons to insert randomly into the genome of a microorganism, under permissive conditions. The transposons are individually marked for easy identification, and then introduced separately into a microorganism, resulting in disruption of the genome. Mutated microorganisms with reduced virulence are then detected by negative selection and the genes where insertional inactivation has occurred are identified and characterised.

A first stage in the STM process is the preparation of suitable transposons or transposon-like elements. A library of different transposons are prepared, each being incorporated into a vector or plasmid to facilitate transfer into the microorganism. The preparation of vectors with suitable transposons will be apparent to a skilled person in the art and is further disclosed in WO-A-96/17951. For the Gram-negative bacteria, e.g. *E. coli*, suitable transposons include Tn5 and Tn10. Having prepared the transposons, mutagenesis of a bacterial strain is then carried out to create a library of individually mutated bacteria.

Pools of the mutated microorganisms are then introduced into a suitable host. After a suitable length of time, the microorganisms are recovered from the host and those microorganisms that have survived in the host are identified, thereby also identifying the mutated strains that failed to survive, i.e. avirulent strains. Corresponding avirulent strains in a stored library are then us d to identify the genes where insertional inactivation occurred. Usually, the site of

transposon insertion is identified by isolating the DNA flanking the transposons insertion site, and this permits characterisation of the genes implicated in virulence.

Once an avirulent microorganism has been identified, it is possible to determine more fully the potential role of the mutated gene in virulence, by infecting a suitable host animal with a lethal dose of the mutant. The survival time of the infected animal is compared with that of a control infected with the wild-type strain, and those animals surviving for longer periods than the control may be said to be infected with microorganisms having mutated virulence genes.

Alternatively, the potential role in virulence can be investigated by infecting an animal host with a mixture of the wild-type and mutant bacteria. After a suitable period of time, bacteria are harvested from organs of the host animal and the ratio of wild-type and mutant bacteria determined. This ratio is divided by the ratio of mutant to wild-type bacteria in the inoculum, to determine the competitive index (CI). Mutants which have a competitive index of less than 1 may be said to be avirulent.

It is possible that the gene which is inactivated by the insertion of the transposon may not be a true virulence gene, but may be having a polar effect on a downstream (virulence) gene. This can be determined by further experimentation, placing non-polar mutations in more defined regions of the gene, or mutating other adjacent genes, and establishing whether or not the mutant is avirulent.

Having characterised a virulence gene in *E. coli*, it is possible to use the gene sequence to establish homologies in other microorganisms. In this way it is possible to determine whether other microorganisms have similar virulence determinants. Sequence homologies may be established by searching in existing databases, e.g. EMBL or Genbank.

Virulence genes are often clustered together in distinct chromosomal regions called pathogenicity islands. Pathogenicity islands can be recognised as they are usually flanked by repeat sequences, insertion elements or tRNA genes. Also the G+C content is normally different from the remainder of the

chromosome, suggesting that they were acquired by horizontal transmission from another organism. For example the G+C content of the *E. coli* K12 genome is 52%. Any pathogenicity islands found in *E. coli* strains are likely to have a G+C content that varies from this average.

The identified virulence genes are likely to be useful both in generating attenuated vaccine strains and as a target for antimicrobials. The same may be true for homologues in Gram-negative bacteria in general.

For the purpose of this invention, the appropriate degree of homology is typically at least 30%, preferably at least 50%, 60% or 70%, and more preferably at least 80% or 90% (at the amino acid or nucleotide level).

Proteins according to the invention may be purified and isolated by methods known in the art. In particular, having identified the gene sequence, it will be possible to use recombinant techniques to express the genes in a suitable host. Active fragments and homologues can be identified and may be useful in therapy. For example, the proteins or their active fragments may be used as antigenic determinants in a vaccine, to elicit an immune response. They may also be used in the preparation of antibodies, for passive immunisation, or diagnostic applications. Suitable antibodies include monoclonal antibodies, or fragments thereof, including single chain fv fragments. Methods for the preparation of antibodies will be apparent to those skilled in the art.

The preparation of vaccines based on attenuated microorganisms is known to those skilled in the art. Vaccine compositions can be formulated with suitable carriers or adjuvants, e.g. alum, as necessary or desired, and used in therapy, to provide effective immunisation against *E. coli* or other Gram-negative bacteria. The preparation of vaccine formulations will be apparent to the skilled person.

More generally, and as is well known to those skilled in the art, a suitable amount of an active component of the invention can be selected, for therapeutic use, as can suitable carriers or excipients, and routes of administration. These factors will be chosen or determined according to known criteria such as the

nature/severity of the condition to be treated, the type or health of the subject etc.

The following Examples illustrate the invention. For the Examples, STM was used to screen an *E. coli* K1 mini-Tn5 mutant bank for attenuated mutants, using a mouse model of systemic infection. The basic procedure followed that disclosed in Hensel *et al*, *supra*. *E. coli* K1 containing a mini-Tn5 insertion within a virulence gene was not recovered from mice inoculated with a mixed population of mutants, and is therefore likely to be attenuated.

The DNA region flanking either side of the mini-Tn5 insertion was cloned by inverse PCR or by rescue of a kanamycin-resistance marker. In the latter case, chromosomal DNA from the STM-derived mutant was digested with restriction enzymes, ligated into the plasmid pUC19, and kanamycin-resistant clones selected after transformation into competent *E. coli* K12 cells. Subsequent cloning and sequencing was then performed and the gene sequences compared using sequences in publicly available sequence databases (EMBL) to help characterise the putative gene products.

<u>Example 1</u>

In a first mutant, two fragments of cloned DNA were sequenced. The nucleotide sequences are shown as SEQ ID NO. 1 and SEQ ID NO. 3 and a translated region of the DNA from SEQ ID NO. 1 is shown as SEQ ID NO. 2. SEQ ID NO. 1 shows 99.8% identity to the *mdoGH* region from *E. coli* K12 (EMBL database accession number AE000206) from nucleotides 2577 to 6908. This DNA fragment encodes the 5'-part of the *ymdD* gene, the entire *mdoG* gene and the 5'-part of the *mdoH* gene. The product of the *mdoG* gene is of unknown function, but is believed to be involved in the biosynthesis of membrane-derived oligosaccharides.

SEQ ID NO. 3 shows 98.3% identity to the 3'-part of the *mdoH* gene and downstream gene sequences from *E. coli* K12 (nucleotides 7187 to 7760). SEQ ID NO. 2 shows 99.6% identity to the *mdoG* protein from *E. coli* K12 (Swiss Prot accession number P33136) at amino acid 1 to 511.

The nov I gene was tested for attenuation of virulence, using mixed infections, in a murine model of systemic infection (Achtman *et al.*, Infection and

Immunity, 1983; Vol. 39:315-335), and shown to be attenuated with a competitive index (CI) of 0.38. This confirms that the attenuation of the original transposon mutant is likely to be due to the disruption of the *mdoG* gene.

Polar and a non-polar deletion mutants of *mdoG* were constructed. The *mdoG* gene and flanking regions were amplified by PCR with oligonucleotides 5'-TGCTCTAGAGCCATTACTCAGAATGGG-3' (SEQ ID NO. 49) and 5'-CGCGAGCTCGACGACTGAATGATCCC-3' (SEQ ID NO. 50). The product was cloned into pUC19. A PCR product containing 5'- and 3'-terminal fragments of *mdoG* and the entire pUC19 sequence was then amplified by inverse PCR with the oligonucleotides 5'-TCCCCCGGGTACTGCAGCACTCAACC-3' (SEQ ID NO. 51) and 5'-GATCCCGGGACCACTGAAATGCGTGC-3' (SEQ ID NO. 52). A non-polar kanamycin resistance cassette (*aphT*) was inserted in both orientations between the *mdoG* sequences to give a polar and a non-polar construct. The *mdoG::aphT* fusions were then transferred to the suicide vector pCDV442. The chromosomal copy of the *mdoG* was mutated by allelic transfer after conjugation of the pCDV442 constructs into wild type *E. coli* K1.

The contructed mutants were tested for attenuation of virulence in a murine model of systemic infection (Achtman et al., *supra)*. Both the polar and the non-polar constructs were attenuated in virulence, with competitive indices of 0.37 and 0.35, respectively (mean CI from three mice each). This confirms that the attenuation of the original transposon mutant is likely to be due to the disruption of the *mdoG* gene.

<u>Example 2</u>

A second mutant was identified with a virulence gene having the nucleotide sequence shown in SEQ ID NO. 4 and the translated amino acid sequence shown as SEQ ID NO. 5. The mini-Tn5 transposon inserted at nucleotide 581 (SEQ ID NO. 4) and at amino acid 187 (SEQ ID NO. 5).

These sequences show 97.9% identity to the *creC* gene of *E. coli* K12 (EMBL and Genbank accession numbers M13608, AE000510 and U14003).

The *creC* protein from *E. coli* K12 belongs to the protein family of histidine kinases as well as to a protein family consisting of proteins containing a signal domain.

The novel gene was tested for attenuation of virulence (Achtman *et al*, *supra*.), and shown to be attenuated with a competitive index of 0.09.

As the *E. coli* K12 *creC* gene is transcribed as part of an operon with the *creD* gene, it is possible that this attenuation is due to a polar effect on a presumed *E. coli* K1 *creD* gene.

<u>Example 3</u>

A third mutant had a nucleotide sequence shown as SEQ ID NO. 6 immediately following the mini-Tn5. A translation of this sequence is shown as SEQ ID NO. 7.

The nucleotide sequence shows 93.7% identity to the *recG* gene of *E. coli* K12, at nucleotides 5-146 (EMBL and Genbank accession numbers P24230 and M64367). This demonstrates that the disrupted gene is at least partially identical to the *recG* gene of *E. coli* K12. The *recG* gene of *E. coli* K12 encodes a 76.4kD protein which functions as ATP-dependent DNA helicase, and plays a critical role in DNA repair.

In tests for attenuation, the competitive index was shown to be 0.48. The *recG* gene is transcribed as the terminal gene of an operon, and it is therefore unlikely that this attenuation is due to a polar effect on another *E. coli* K1 gene.

<u>Example 4</u>

A fourth mutant had a transposon inserted within the nucleotide sequence shown as SEQ ID NO. 8, with a translation product shown as SEQ ID NO. 9.

The mini-Tn5 transposon inserted at nucleotide 359 and amino acid 80.

These sequences show 98.5% sequence identity to the *yggN* gene of *E. coli* K12 (EMBL accession number AE000378) at nucleotides 339-1054, and 99.6% identity at the amino acid level.

Although the sequence of the *yggN* gene is known, the function of its encoded protein has not yet been determined.

The novel gene was tested for attenuation of virulence, and shown to be attenuated with a competitive index of 0.43.

<u>Example 5</u>

Several mutants were also found with a transposon insertion within the same region. Cloning and sequencing the region revealed a nucleotide sequence shown as SEQ ID NO. 10. This sequence has homology with the *tatABCD* operon of *E. coli* K12 (EMBL and Genbank accession numbers AJ005830, AE000459 and AE000167). This operon encodes proteins of predicted mass 9.6 kD, 18.4 kD, 28.9 kD and 29.5 kD, which function as components of a Sec-independent protein export pathway. The pathway permits translocation of fully folded proteins to the periplasm through a gated pore, after the attachment of co-factors in the cytoplasm.

Translation of the nucleotide sequence revealed a protein corresponding to *tatA* (SEQ ID NO. 11), a sequence corresponding to *tatB* (SEQ ID NO. 12), a sequence corresponding to *tatC* (SEQ ID NO. 13) and a sequence corresponding to *tatD* (SEQ ID NO. 14).

The mini-Tn5 transposons in the mutants identified by STM are located at nucleotides 1429 and 2226 of SEQ ID NO. 10. These transposon insertions disrupt the *tatB* protein sequence at amino acid 50 and the *tatC* protein sequence at amino acid 143.

The *tatB* and *tatC* genes were tested for attenuation of virulence and were shown to be attenuated with competitive indices of 0.0012 and 0.0039, respectively. These genes were also attenuated in virulence when tested in single infections in the same model of systemic infection.

<u>Example 6</u>

A further mutant was insertionally inactivated within a region corresponding to the *tatE* gene of *E. coli* K12, shown as SEQ ID NO. 15. A translation of the sequence as shown as SEQ ID NO. 16. The *tatE* gene shows 98% identity to that of the *E. coli* K12 gene (accession number AE000167) at nucleotides 6719-7306.

To establish whether the *tatA, tatD* and *tatE* genes are required for virulence, non-polar deletion mutations were constructed in each. The regions of DNA flanking either side of the *tatA*, *tatD* and *tatE* genes were amplified with the following primers:

tatA

5'-TCG TCT AGA GAT GAT GGT GAT GGA GCG-3' (SEQ ID NO. 53)

5'-GAA CTG CAG CCA AAT ACT GAT ACC ACC C-3' (SEQ ID NO. 54)

5'-GAA CTG CAG GCT AAA ACA GAA GAC GCG-3' (SEQ ID NO. 55)

5'-CAT GCA TGC ACT CCA TAT GAC AAC CGC-3' (SEQ ID NO. 56)

Primers SEQ ID NO. 53 and SEQ ID NO. 54 were used to amplify DNA sequences upstream of *tatA*, Primers SEQ ID NO. 55 and SEQ ID NO. 56 were used to amplify DNA sequences downstream of *tatA*.

tatD

5'-TCG TCT AGA ATG AAG CTG CGC ATG AGG-3' (SEQ ID NO. 57)

5'-CAA CTG CAG TCG CAA ATT GCG AAC TGG-3' (SEQ ID NO. 58)

5'-CAA CTG CAG ACC GCA ACT TTT CGA CGC-3' (SEQ ID NO. 59)

5'-CAT GCA TGC CAG TGA GCC ATT GTT CCC-3' (SEQ ID NO. 60)

Primers SEQ ID NO. 57 and SEQ ID NO. 58 were used to amplify DNA sequences upstream of *tatD*, Primers SEQ ID NO. 59 and SEQ ID NO. 60 were used to amplify DNA sequences downstream of *tatD*.

tatE

5'-TGC TCT AGA TAC GAC TCT GAC AGG AGG-3' (SEQ ID NO. 61)

5'-TCA GAT ATC AAC TAC CAG CAG TTT GG-3' (SEQ ID NO. 62)

5'-TCA GAT ATC CAT AAA GAG TGA CGT GGC-3' (SEQ ID NO. 63)

5'-TGC TCT AGA AAA CGT GGC AAC AGA GCG-3' (SEQ ID NO. 64)

Primers SEQ ID NO. 61 and SEQ ID NO. 62 were used to amplify DNA sequences upstream of *tatE*, Primers SEQ ID NO. 63 and SEQ ID NO. 64 were used to amplify DNA sequences downstream of *tatE*.

After cloning these flanking DNA fragments into pUC19, a non-polar *aphT* kanamycin resistance cassette (Galan *et al*, J.Bacteriol, 1992;174:4338-4349) was inserted between the flanking DNA fragments to replace the *tatA*, *tatD* and *tatE* genes. These DNA fragments were then transferred to the suicide vector pCVD442 (Blomfield *et. al*, Mol. Micro., 1991;5:1447-1457). The chromosomal copies of the *E. coli* K1 *tatA*, *tatD* and *tatE* genes were then mutated by allelic transfer after conjugation of the pCVD442 constructs into wild type *E. coli* K1.

Disruptions of the *tatA*, *tatD* and *tatE* genes have been tested for attenuation of virulence (Achtman *et al*., *supra*).

None of the genes was attenuated when deleted in isolation. The genes may still play a role in virulence, and to test this, mutants were prepared with deletions in both *tatA* and *tatE* genes. The double mutant was tested for attenuation in virulence using mixed infections with the wild-type strain and shown to be attenuated with a competitive index of 0.0017. It seems therefore that the *tatA, tatD* and *tatE* genes may be used in combination to create avirulent microorganisms.

Given the similarity of the *E. coli* K1 *tatABCD* genes to predicted *tatABCD* genes present in the *S. typhimurium* genome and *Neisseria meningitidis* genome it seemed likely that the tat system may also be required for virulence in these, and other, organisms. A deletion in the *S. typhimurium tatC* gene (SEQ ID NO. 17) was constructed by amplifying the DNA flanking either side of the *tatC* gene with the following primers:

5'-TGC TCT AGA AGG CGT TGT CGA TCC TG-3' (SEQ ID NO. 65)

5'-GAA CTG CAG GAA AAG GCC GAG CAG ACT G-3' (SEQ ID NO. 66)

5'-GAA CTG CAG TAC AGC CAT GTT TAC GGT-3' (SEQ ID NO. 67)

5'-CAT GCA TGC GGT GTA CGA CAG TTT GCG-3' (SEQ ID NO. 68)

Primers SEQ ID NO. 65 and SEQ ID NO. 66 were used to amplify DNA sequences downstream of the *S. typhimurium tatC* gene, Primers SEQ ID NO. 67 and SEQ ID NO. 68 were used to amplify DNA sequences upstream of the *S. typhimurium tatC* gene.

The encoded amino acid sequences for two regions of the tatC gene are shown as SEQ ID NO. 18 and SEQ ID NO. 19.

After cloning these flanking DNA fragments into pUC19, a non-polar kanamycin resistance cassette (*aphT*) was inserted between the flanking DNA fragments to replace the *S. typhimurium tatC* gene. This DNA fragment was then transferred to the suicide vector pCVD442. The chromosomal copy of the *S. typhimurium tatC* gene was then mutated by allelic transfer after conjugation of the pCVD442 construct into wild type *S. typhimurium* strains TML and SL1344.

The disrupted *S. typhimurium tatC* gene was tested for attenuation of virulence, using mixed and single infections in a murine model of systemic infection. For mixed infections, 6-7 week old *balbC* mice were inoculated intraperitoneally with 10^4 bacterial cells. Competitive indices were calculated after comparing the numbers of mutant and wild-type bacteria present in spleens after 3 days. For single infections, mice were inoculated either intraperitoneally or orally with varying doses and mouse survival monitored for 17 days. The strains were attenuated in virulence, the competitive indices of the SL1344 *tatC* and TML *tatC* deletion strains being 0.078 and 0.098, respectively.

In single infections, mouse survival was extended compared to the wild-type controls.

Sequence homology was also demonstrated with the tat sequence from *Neisseria meningitidis*. The gene sequence from *N. meningitidis* is shown as SEQ ID NO. 20 and the encoded amino acid sequence for tatC is shown as SEQ ID NO. 21.

To test for virulence, a deletion mutant was created using the following primers:

5'-TGCTCTAGACACATCATGGCACACC-3' (SEQ ID NO. 69)

5'-GAACTGCAGAACCGTCCACATCAGGCG-3' (SEQ ID NO. 70)

5'-GAACTGCAGACCCTGCTTGCCATTCCG-3' (SEQ ID NO. 71)

5'-GAACTGCAGACCCTGCTTGCCATTCCG-3' (SEQ ID NO. 72)

Cloning of the DNA fragments and the *aphT* kanamycin resistance cassette into pUC19 followed the procedure outlined above for *S. typhimurium*. The chromosomal copy of the *N. meningitidis tatC* gene was mutated by transformation of the pUC19-based constructs into wild-type *N. meningitidis* cells.

Southern analysis of the resulting transformants indicated that all the transformants were merodiploids and contained both the wild-type and mutated copies of the *tatC* gene. This indicates that there is some selection against the isolation of mutants in which the *tatC* gene has been deleted.

Further studies on polar and non-polar constructs showed that transformants did not grow on selective media. This suggests that the *N. meningitidis tatC* gene is essential for the *in vitro* growth of this organism.

<u>Example 7</u>

A further mutant was identified with a transposon insertion within a nucleotide sequence identified herein as SEQ ID NO. 22, at nucleotide 3981. The sequence defined herein as *eck1*, shows sequence homology to several Group 1 glycosyltransferases from a number of bacteria. Sequence homology was also shown to the *gnd* gene of *E. coli* K12 (at nucleotides 4197-4604 of SEQ ID NO. 22).

The translation of the *E. coli eck1* gene is shown as SEQ ID NO. 26. The gene has been tested for attenuation of virulence, as described above, and is shown to be attenuated with a competitive index of 0.025.

Several open reading frames (ORF) were also identified from the DNA sequence (SEQ ID NO. 22). The first of these is defined herein as MS1 and a translation product shown as SEQ ID NO. 25. The amino acid sequence is shown to have 50.3% identity to a putative glycosyl transferase from *E. coli*

serotype 0111 (TrEMBL database accession number AAD46732). The amino acid sequence also shows homology with the eck1 protein from *E. coli* K1 and also the TrsE protein from *Yersinia entercolitica* (TrEMBL database accession number Q56917).

A second open reading frame identified herein as MS2 had the gene sequence shown as SEQ ID NO. 24. This shows sequence homology to the putative glycosyl transferase TrsC from *Yersinia entercolitica* (TrRMBL database accession number Q56915), and also the glycosyl transferase WbnA from *E. coli* serotype 0113 (TrEMBL database accession number AAD50485).

A third open reading frame encodes a product identified herein as MS3 (SEQ ID NO. 23). The amino acid sequence shows 30.2% identity to a rhamnosyltransferase from *Streptoccus mutans*.

The gene sequence shown as SEQ ID NO. 22 may be at least part of a pathogenicity island, with multiple virulence genes being positioned in a cluster on the microorganism's genome.

<u>Example 8</u>

A further mutant was identified having a transposon insertion within the *iroCDE* operon. The nucleotide sequences flanking either side of the mini-Tn5 insertion are shown as SEQ ID NO. 27 and SEQ ID NO. 30.

The mini-Tn5 transposon is inserted at nucleotide 1272 of SEQ ID NO. 27 and at nucleotide 1 of SEQ ID NO. 30, and interrupts the *iroD* gene. The N-terminal region of *iroD* is shown as SEQ ID NO. 29, and the C-terminal region is shown as SEQ ID NO. 31.

In addition to *iroD*, the gene shown as SEQ ID NO. 27 encodes a partial peptide with the amino acid sequence shown as SEQ ID NO. 28. This amino acid sequence shows 70.9% identity to the putative ATP binding cassette transporter *iroC* from *Salmonella typhi*.

The gene sequence shown as SEQ ID NO. 30 includes an open reading frame that encodes a peptide with the amino acid sequence shown as SEQ ID NO. 32 and this has sequence homology to the *iroE* protein from *Salmonella typhi*.

Testing the genes in a model for attenuation of virulence, as described above, showed that the *iroD* gene was attenuated with a competitive index of 0.107. The mini-Tn5 mutation in the iroD gene has been reintroduced into the wild-type *E. coli* K1 strain by P1 transduction. The resulting transductant is also attenuated in virulence with a competitive index of 0.1. This indicates that the attenuated phenotype is linked to the insertion within *iroD*. However, it is possible that the attenuation is due to a polar effect on the *E. coli* K1 *iroE* gene.

Example 9

A further mutant was identified with a transposon insertion within the nucleotide sequence shown as SEQ ID NO. 33. The transposon is inserted at nucleotide 2264 of SEQ ID NO. 33. The nucleotide sequence shows sequence homology to the *aslA / hemY* region of *E. coli* K12 (EMBL accession number AE000456). The *aslA* encodes an arylsulfatase homologue whereas *hemY* is involved in the biosynthesis of protoheme IX. This demonstrates that the disrupted region is at least partially identical to the *aslA / hemY* region of *E. coli* K12.

The transposon is inserted at nucleotide 2264 of SEQ ID NO. 33. This insertion site is 216 nucleotides downstream from the stop codon of the *hem*Y gene and 472 nucleotides upstream from the start codon of the *aslA* gene.

The novel region has been tested for attenuation of virulence, as described above, and shown to be attenuated with a competitive index of 0.033. The mini-Tn5 mutation in this region has been reintroduced into the wild-type *E. coli* K1 strain by P1 transduction. The resulting transductant is also attenuated in virulence with a competitive index of 0.008. This indicates that the attenuated phenotype is linked to the transposon insertion in this region. However, polar and non-polar deletion mutants of *aslA* were constructed and tested for attenuation of virulence as described above.

Neither the polar nor the non-polar mutants were attenuated in virulence and this demonstrates that the attenuation of the original transposon mutant is not due to a polar effect on the *asl*A gene. This indicates that the transposon is disrupting some other function ncoded within the intergenic region between *asl*A and *hem*Y. For example ther could be some untranslated RNA molecule,

such as a regulatory RNA similar to oxyS (Altuvia *et al.*, Cell, 1997;90:43-53), encoded within this region. Alternatively the transposon could be disrupting some DNA structure that may, for example, be involved in DNA replication. This DNA region is also present in the pathogen *Salmonella typhimurium* suggesting that it may be important for pathogenicity in other organisms. This region (SEQ ID NO. 33) may be used as a target, to identify anti-microbial drugs.

Example 10

A further mutant was identified and the DNA region flanking either side of the mini-Tn5 insertion was cloned and had the nucleotide sequence shown as SEQ ID NO. 34. This nucleotide sequence has homology with the *mtd2* gene of *Herpetosiphon aurantiacus* (EMBL accession number P25265), with the *mtd2* gene product functioning as a cytosine-specific methyltransferase. The *mtd2* gene is not found in the *E. coli* K12 genome and may represent a pathogenicity island.

The mini-Tn5 transposon insertions were located at nucleotides 4773 and 3764 of SEQ ID NO. 34 and were shown to interrupt the *mtd2* gene.

The amino acid sequence of the *mtd2* gene is shown as SEQ ID NO. 43.

The *E. coli* K1 *mtd2* gene was tested for attenuation of virulence, as described above, and shown to be attenuated with a competitive index of 0.073.

In addition to the *mtd2* gene, a series of open reading frames were also identified with translation products identified herein as MS4 to MS16, SEQ ID NOS. 48-44 and 42-35, respectively. As the open reading frames are located in a potential pathogenicity island, mutations in these genes may also result in attenuation in virulence. Further, since it is known that *E. coli* and other bacteria may encode peptides in different forms in the nucleotide sequence, the coding regions of some of these proteins may overlap. In addition, any aminoacid sequence shown starting with Val may in fact start with Met.